AFRL-RH-WP-TR-2010-0130

AUTOMATED SENTIMENT ANALYSIS

Stephen M. Shellman
Strategic Analysis Enterprises, Inc.
108 Bluffs Circle
Williamsburg VA 23185

Michael A. Covington
The University of Georgia
230 River Road
Athens GA 30692

JUNE 2009
Final Report

Approved for public release; distribution is unlimited.

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2010-0130 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//
LAURIE H. FENSTERMACHER
Work Unit Manager
Behavior Modeling Branch

//SIGNED//
GLENN W. HARSHBERGER
Anticipate & Influence Behavior Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
24-06-2009	Final	August 2008 – June 2009

4. TITLE AND SUBTITLE

Automated Sentiment Analysis

5a. CONTRACT NUMBER FA8650-08-C-7864

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER 63123F

6. AUTHOR(S)

[1]Stephen M. Shellman, [2]Michael A. Covington

5d. PROJECT NUMBER DRPA

5e. TASK NUMBER 05

5f. WORK UNIT NUMBER DRPA0519

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

[1]Strategic Analysis Enterprises, Inc.
108 Bluffs Circle
Williamsburg VA 23185

[2] The University of Georgia
230 River Road
Athens GA 30602

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Anticipate & Influence Behavior Division
Behavior Modeling Branch
Wright-Patterson AFB OH 45433-7022

10. SPONSOR/MONITOR'S ACRONYM(S)

711 HPW/RHXB

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-RH-WP-TR-2010-0130

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

88ABW/PA cleared on 10 Nov 10, 88ABW-2010-6008.

14. ABSTRACT

This project develops a software tool to automate the collection of sentiment data from electronic media reports. Through analyses the report illustrates that the data generated from the new software tool produces valid and reliable data measuring attitudes and beliefs about political candidates, their actions, and their policies. The report further shows that such data improves the model fit of political behavior models. Specifically, the new data reveals important impacts on the actions and interactions of political actors. The report closes with the prospects for future analyses and research in this area and the area of national security.

15. SUBJECT TERMS

Sentiment Analysis, Communication, Illocution, Linguistic Modeling, Location, Perlocution

16. SECURITY CLASSIFICATION OF: UNCLASSIFED			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Laurie H. Fenstermacher
c. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	SAR	50	NA

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE LEFT INTENTIONALLY BLANK

TABLE OF CONTENTS

Section	Page
1.0 Introduction…………………………………………………………………………	1
2.0 Goals and Accomplishments……………………………………………………......	3
3.0 The Current Sentiment Literature……………………………………………………..	6
3.1 What is Sentiment Analysis…………………………………………………...	6
3.2 Issues in Sentiment Analysis…………………………………………………	6
3.3 Theoretical Approach to Sentiment and Political Events Analysis…………...	7
3.4 Key Literature…………………………………………………………………	8
4.0 Our Approach to Automating Sentiment……………………………………………...	10
4.1 Pathos………………………………………………………………………….	10
4.1.1 Text Classification…………………………………………………......	10
4.1.2 BoWs…………………………………………………………………..	11
4.1.3 Utterances & Speech Acts……………………………………………..	11
5.0 Research Design……………………………………………………………………...	14
5.1 Data…………………………………………………………………………….	14
5.1.1 News Sources……………………………………………………………..	14
5.1.2 Blog Sources……………………………………………………………...	14
5.1.3 Polls…………………………………………………………………….....	14
5.1.4 Events Data……………………………………………………………....	14
5.1.5 Other Country-Level Data…………………………………….........	15
5.1.6 Data Aggregation…………………………………………………...	15
5.2 Methodology…………………………………………………………………	15
5.2.1 Internal Validity Methodology……………………………………....	15
5.2.2 External Validity Methodology……………………………………..	15
6.0 Results………………………………………………………………………………...	17
6.1 Interpreting the Graphs………………………………………………………..	17
6.2 Ma Monthly Results………………………………………………………..	17
6.3 Ma Weekly Results…………………………………………………………	24
6.4 Monthly Economic Results……………………………………………………	27
6.5 Weekly Security Results………………………………………………………	31
6.6 DARPA SAE ICEWS Rebellion Model with Sentiment as an Added Factor..	35
7.0 Future Technical Improvements……………………………………………………...	37
7.1 SAE Internal Research and Development (IRAD)……………………………	37
7.1.1 Verb Dictionary………………………………………………………..	37
7.1.2 Actor Dictionaries……………………………………………………...	37
7.1.3 Pathos Improvements…………………………………………………	37
8.0 Future Projects……………………………………………………………………....	39
8.1 Strategic Communications…………………………………………………….	39
8.2 Effects Based Operations……………………………………………………....	39
9.0 Conclusion…………………………………………………………………………….	40
References……………………………………………………………………………….	41
List of Acronyms………………………………………………………………………..	42
Glossary…………………………………………………………………………………	43

LIST OF FIGURES

Figure	**Page**
Figure 1: Model of Attitudes & Events………………………………………………..	4
Figure 2: Pathos Software Tool………………………………………………………..	10
Figure 3: Comparisons of Monthly Ma Series……………………………………....	18
Figure 4: Model of Polls Using Our Sentiment Measures……………………….	19
Figure 5: VAR Models of Ma Hostile Actions…………………………………..	21
Figure 6: VAR Models of Ma Cooperative Actions……………………………..	22
Figure 7: VAR Models of Sentiment…………………………………………….	23
Figure 8: Comparison of Weekly Ma Series……………………………………..	24
Figure 9: VAR Models of Ma Hostile Actions (Weekly)………………………..	26
Figure 10: Out of Sample Forecasting (10 weeks)………………………………	27
Figure 11: Comparisons of Monthly Economic Series…………………………..	28
Figure 12: VAR Models of Ma Hostile Actions (Weekly)………………………	30
Figure 13: Comparisons of Weekly Security Series……………………………..	31
Figure 14: VAR Models of Taiwan-China Relations (Weekly)…………………	33
Figure 15: Out of Sample Forecast from Full VAR Foreign Policy Model: Taiwan Cooperation Toward China (20 weeks)………………………	34

LIST OF TABLES

Table	**Page**
Table 1: Breakdown Across Media & Blogs……………………………………....	18
Table 2: Correlating Ma Sentiment Measures to Economic Measures……………	19
Table 3: Breakdown Across Media & Blogs……………………………………....	24
Table 4: Correlating Economic Sentiment Measures to Economic Measures…….	29
Table 5: Breakdown of Unification Across Media & Blogs (Weekly)……………	32
Table 6: ICEWS Rebellion Model with Sentiment Included……………………..	35

THIS PAGE LEFT INTENTIONALLY BLANK

1.0 INTRODUCTION

> *"We must rely on the force of the popular masses, for it is only thus that we can have a guarantee of success."* (Tse-Tung, M., 1966, 57)
>
> *"the guerrilla fighter needs full help from the people of the area. This is an indispensable condition…" and guerrillas must draw their "greatest force from the mass of the people."* (Guevara, C., 1985, 50)
>
> *If she "loses the loyalty of a sufficient number of members of the winning coalition, a challenger can remove or replace her in office."* (Bueno de Mesquita, B., et al., 1999)

Our project developed an enabling technology that validly and reliably generates data to measure sentiment in more efficient and effective ways. Automating sentiment analysis can pay huge dividends in aiding our understanding of political dynamics, strategic communications, and effects based operations – better, faster, & cheaper. In this seedling, we showed that our newly generated sentiment data closely mirrors polling data and performs well in models of politics increasing various models' explanatory power. That is, our models which include sentiment better explain and predict political behavior with less error than models which exclude such data. In short, our new data outperforms polling data given that polling data is costly to collect, contains error, and is difficult to acquire in real time for various parts of the world. Moreover, we demonstrate that sentiment is an important variable to include in models of politics, and without it, models are plagued with omitted variable bias.

As the quotes from Tse-Tung, Guevara, and Bueno de Mesquita, et al., above indicate we know that support from the masses impacts political violence and politics more broadly. Yet, empirical studies are limited by a dearth of data to test how policies, actions and personalities shape attitudes and beliefs and how such attitudes and beliefs effect various actors' strategies, tactics, and policies. Traditionally, polling data was the only way to measure and include such indicators in models of politics. However, polls are infrequent, expensive, and complicated to carry-out in certain locations. As a result, sentiment is difficult to measure in near real time and across space (cities, towns, regions, countries, etc.).

Advances in linguistics and technology allowed us to develop a software program capable of automating the collection of sentiment across space and time. Specifically, the application of semantic analysis, in particular Discourse Representation Theory (DRT), together with a syntactic parser measures positive and negative opinions and attitudes in documents, news reports, blogs, and websites, which we then roll up into a measure of sentiment. We test the accuracy of our measures by comparing them to sparsely available public opinion data.

Upon showing that our new measures are reliable and valid indicators of sentiment, we implement our new, near-real time measures of sentiment in models of politics. Our results show that sentiment aids in explaining and forecasting indicators of political behavior. The models which include sentiment outperform (i.e., provide better fit than) the models which exclude our measures.

Our report progresses by first laying out our project's goals and accomplishments. Second, we review the current literature and state of the art techniques. Third, we explain our approach and how it differs from and advances the current approaches. Fourth, we describe our new software tool which we call *Pathos*. Fifth, we walk the reader through the data generation process. Sixth, we outline our research design and how we validate our measures as well as how we determine whether or not our new measures further our understanding of politics and economics. Seventh, we communicate our results. Finally, we close by discussing the possibilities that our new tool gives rise to and conclude with some brief remarks.

2.0 GOALS AND ACCOMPLISHMENTS

The technical research challenge is to provide data on near-real time attitudes, opinions, and beliefs about politics. To do so, we must contrive a software program to read in near-real time online text sources (e.g., blogs and media reports) and calculate various measures of sentiment. The big question is not necessarily can we generate a program to code text. Rather, we ask, can we generate a program that will produce measures of sentiment that are representative of various actors' (and the larger population's) attitudes and beliefs. To accomplish this task, we will bring together political event analysis, linguistics, event understanding, and advances in artificial intelligence.

We define sentiment as attitudes and opinions about a phenomenon. Most choose to measure such a concept using subject matter experts (SMEs) or public opinion data. SMEs can be inaccurate and provide subjective assessments of an actor's or populations' attitudes. Public opinion can often be very time consuming and take years and months to develop instruments, administer the surveys to the target actors or population, and collect the desired data. Such an approach yields measures of sentiment at a particular point in time and is difficult to produce over multiple time points (i.e., produce data over small repeatable time horizons). Most polling data are static with few exceptions where firms, political climates, and resources merge to allow multiple polls to be administered daily or weekly (e.g., advanced industrial democracies). However, such perfectly aligned cases are not the norm and attempting to get daily polling data in less developed, conflict-ridden regions is most difficult.

The current state of sentiment analysis is in its infancy stages and is prone to many drawbacks. First, the majority of sentiment analysis focuses on how the writer feels about such phenomenon (e.g., how a reviewer or critic feels about a movie). In political research, Social Science Automation (SSA) has concentrated on developing a tool to analyze how an author from a dissident group website (e.g., Hamas) perceives politics. They use a bag of words (BoWs) technique to calculate the number of positive and negative words appearing in a specific story, blog posting, thread, or forum. While this is useful information, the sentiment is not attributable to a specific person and is difficult to tie to specific policies and actions. In short, the actor responsible for the sentiment and the target of the sentiment is often missing. If we want to know how certain types of people feel about a government policy or action, we cannot directly glean this information from such a strategy. While the approach is useful in understanding a particular group's (e.g., Hamas) opinions, it cannot measure how alternative actors feel about policies, actions, and/or political actors.

Events data contain information about who did what to whom as reported in the open press. The strength of such data is that they contain "objective" information about the actions of one actor towards another actor. This is exactly the structure of information we desire when it comes to understanding various actors' sentiment; we desire to know who is saying what about who and/or what. That is, we desire to code what we refer to as "utterances." An utterance is a complete unit of speech in spoken language. For example, we want to know what various leaders of organizations and groups, social actors, and ordinary citizens are saying about government policies. We can then break these utterances down and analyze them for meaning and interpret them within context.

Sentiment analysis based on vocabulary is well known, however, a more structured kind of sentiment analysis with more understanding of the semantics of the events being described remains absent from the current literature and technologies. Our seedling fills that gap. We take

the BoWs approach as well as attempt to code utterances. We roll both sets of information up into measures of sentiment and compare them to each other, traditional polling data, and additional indicators to examine their validity and reliability. Our focus on utterances and speech acts identifies specific actor's opinions as distinct from the mere intent or stance of the writer, and ties it back to event understanding. Previous work in this area does not relate sentiment to events.

Figure 1 shows a general model of how attitudes can be related to events. We hypothesize that the general public's attitudes have an effect on government and dissident behavior, while government and dissident behavior and the interactions of their behavior impact the general public's attitudes and beliefs. Our project collects information on each of the relationships represented in Figure 1 and tests whether or not our hypothesis is supported. For the seedling, we concentrated on the effects of sentiment on government behavior and the effects of government behavior on sentiment.

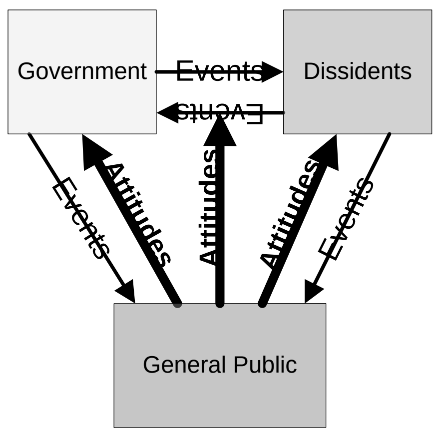


Figure 1: Model of Attitudes & Events

Following the development of our new software tool and the collection of our new data, we performed a rigorous quantitative empirical analysis to examine (1) the validity and reliability of our new indicators and (2) the utility of sentiment data in models of politics such as those inferred from Figure 1. We focused on Taiwan for this analysis after consulting the Defense Advanced Research Projects Agency (DARPA) program manager, Sean O'Brien. When compared to polling measures we were able to collect from Taiwan, our new indicators appeared internally valid and reliable. The new measures also were deemed externally valid when they correlated in the same direction and relative magnitude as the polling measures with other measures sentiment is generally thought to correlate with (e.g., Taiwan economic measures, political performance measures, etc.). Moreover, our statistical models built following our depiction in Figure 1reveal that sentiment is an important variable missing in previous studies of politics and has important effects on government and dissident behavior. The results suggest that automating the collection of sentiment data can increase our understanding of political conflict and improve the accuracy of our forecasting models.

In sum, we developed a prototype software program to validly and reliably measure sentiment in near real-time in two different ways. We then showed our measures are valid and

reliable. Finally, we showed the utility these measures have in models of politics, political conflict, and economics. Below we discuss the literature and technology we built upon and elaborate on our methodology.

3.0 THE CURRENT SENTIMENT LITERATURE

3.1 What is Sentiment Analysis?

Deep philosophical questions could be raised about the nature of sentiment. It is not exactly an emotion – one can choose to support a candidate without liking him – but has an evaluative component, i.e., it is a predisposition or position on the part of the speaker.

For purposes of this study, we will define sentiment as all reflections of support, liking, opposition, or disliking of actors or their actions or proposals.

3.2 Issues in Sentiment Analysis

The purpose of our project is not just to measure sentiment, but to link it to actions (physical or verbal) that affect it. Accordingly, sentiment analysis is linked with political events. The use of newspapers as the source material contrasts with much earlier work based on movie or product reviews:

- Unlike product reviewers, newspapers do not assign a "star rating" (1 to 5 stars) to everything they write about; thus they do not tell us, apart from the text, what their sentiment is.

- Newspapers use simple, direct language; this makes lexical and syntactic analysis easier. It also forestalls misunderstanding; sentences likely to be misclassified (because of unusual style, sarcasm, etc.) are not common.

- Newspapers express sentiment both directly and indirectly. One of the most insidious ways newspapers take positions is simply by choosing what to report and what not to report (e.g., highlighting incidents that reflect well or badly on a particular person).

Much work on sentiment analysis involves machine learning, which is of two kinds:

- In supervised learning, you give the computer examples of inputs of various types, and ask it to induce rules that will enable it to classify more inputs along the same lines.

- In unsupervised learning (e.g., clustering), you give the machine a set of inputs and ask it to classify them, putting similar ones together without knowing in advance what properties will play a role in doing this.

Sentiment analysis is an obvious job for supervised learning, where you have texts with known sentiments and you want to find out how they can be distinguished.

Unfortunately, some sentiment analysis studies seem more interested in validating a machine learning technique than in developing a good sentiment analysis technique. A pitfall of machine learning is hiding the relevant information from the computer or making it excessively hard to get. For example, if a computer is required to analyze sentiment on the basis of vocabulary alone ("BoWs"), with no cues indicating sentence structure, it will never distinguish *Germany invaded Poland* from *Poland invaded Germany*. For political event analysis, that is not satisfactory. Nor should it be satisfactory for understand who is saying what about who or what.

Another pitfall is overtraining. A machine learning system can learn unimportant coincidences that impair its performance later on. For example, suppose that in the training corpus, the phrase “on Tuesday” happens to occur mostly in descriptions of some great calamity. Then the phrase “on Tuesday” would be classified as having negative associations even though it should not. For a less fanciful example, consider “11” in texts of American news from late 2001 – it occurs mostly in texts about “9/11” even though the number 11 itself has no emotional significance. We focus on supervised learning for this prototype; though, we can alter our program to perform unsupervised learning classification if desired.

3.3 Theoretical Approach to Sentiment and Political Event Analysis

We apply research in linguistics on discourse analysis, pragmatics, and speech acts to analyze strategic interactions among governments, dissidents, and the citizenry within countries. Pragmatics is the study of the way the use of language relates to the extra-linguistic context and thereby enables speakers to communicate more than that which is explicitly stated. From a pragmatic point of view, there are three main components of a communication. To begin, *locution* means the semantic or literal significance of the utterance. The second component is *illocution* or the intention of the speaker. The last component of a communication, *perlocution*, refers to how the locution was received by the listener and its subsequent effects.

Our key theoretical insight is that these three dimensions, locution, illocution, and perlocution, apply to political actions and reactions whether or not they use language. We contend that political actions such as calls for policy change, nonviolent protests, government repression, and terrorist attacks all contain three components of communication, a literal meaning, an intended meaning, and an interpreted meaning and/or effect. When these three components are out of balance with one another, miscommunication can occur. Its occurrence can yield unexpected and unintended actions and consequences. Our goal for this seedling is to be able to explain and predict perlocutions (events) from locutions (utterances/speech acts). Extensions and future analyses will derive meaning and intended and unintended effects (illocutions) from such communications and examine the balance among locution, illocution, and perlocution and the repercussions when they are out of balance with one another.

We ground our framework in relevant social science and linguistics theories to better conceptualize our three-dimensional analysis of effects based operations. Our *three-dimensional framework* leads to practical computer models and software tools for understanding the intended and unintended consequences of political events. Locutions, the actions themselves, are directly observable and there are well-known methods for coding them (e.g., Text Analysis by Augmented Replacement Instructions (TABARI) and Pericles, etc.). Perlocutions, or effects, are observable as consequences. Illocutions can be inferred as intended or probable effects, based on regular patterns in the course of events. Following our ability to generate a sentiment analysis program and connect locutions to perlocutions, future analyses will engage illocutions and speak more directly to issues of strategic communication and messaging.

The following questions about the text at hand must be answered when performing sentiment analysis:

(1) Does the text express sentiment? To what extent? Indirectly or directly?
(2) What is the sentiment about?
(3) Is it positive or negative?

(4) Is it the writer's own sentiment; is it the sentiment he is trying to inculcate in the reader (possibly different); is he speaking for a third party?
(5) Does it contain sentiment expressed by one actor about another actor or policy?

3.4 Key Literature

In exploring this topic we built upon some key studies. We briefly recap each of the major insights we considered and built upon as we built our software tool.

A. Boiy, E., Hens, P., Deschacht, K., & Moens, M-F. (2007). Automatic sentiment analysis in on-line text. *Proceedings of the 11th International Conference on Electronic Publishing (ELPUB2007)*, pp. 349-360. August 13-15, Vienna, Austria.

This paper starts with a brief but very focused and useful literature review covering the definition of sentiment, seminal work, and methods of measuring sentiment.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1-135.

This 135-page survey of the field is intended for nonspecialists and is not as densely packed with information as other papers. One of the most useful parts is section 7, which lists publicly available lexical resources and other datasets useful for building sentiment analyzers.

C. Shanahan, J.G., Qu, Y., & Wiebe, J. (2006) *Computing attitude and affect in text: theory and applications.* Dordrecht, Netherlands: Springer.

This book comprises 24 papers which we shall review selectively next month.

D. Biber, D. & Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text: Interdisciplinary Journal for the Study of Discourse*, 9(1), pp. 93-124.

This is an early but definitive study of the expression of **evidentiality** (the speaker's confidence in the information reported) and **affect** (the speaker's emotion toward the information reported). The two are grouped together as "stance" and six stance styles are distinguished:

(1) "Emphatic expression of affect" (personal letters, recommendations, romance novels)

(2) "Faceless stance," marked absence of stance features (press reviews, nonfiction, adventure and mystery stories)

(3) "Interactional evidentiality," much personal indication of certainty vs. doubt (personal conversations, personal letters)

(4) "Expository expression of doubt" (academic prose, press reportage)

(5) “Predictive persuasion,” certainty adjectives and predictive modals (letters of recommendation, mainly)

(6) “Oral-controversial persuasion,” frequent predictions and moderately frequent expressions of certainty and possibility (comprises some examples of all genres)

E. Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning, (*paperback edition, 1971). Champaign, IL: University of Illinois Press.

This was one of the first studies to use factor analysis and multidimensional scaling in psychology. Page 37 gives a set of 50 word pairs (e.g., *wide—narrow*) with loadings on four factors (roughly, “evaluation,” “potency,” “activity,” and residual error).

The loading on the first factor is roughly the extent to which the first word in the pair connotes goodness relative to the second one. For example, *good—bad* has a loading of 0.88, and *bitter—sweet* has a loading of −0.80, negative because the “good” word comes second instead of first.

This is not an all-purpose sentiment lexicon because some of the words have curious loadings (e.g., *yellow—blue* is 0.33 on the good—bad scale).

A later chapter of the book discusses consistency in sentiment. Note that the entire book uses 1957 psychology (raw behaviorism) and 1957 mathematics (very early factor analysis and scaling techniques).

F. Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: determining support or opposition from Congressional floor-debate transcripts. *2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pp. 327-335. July 22-23, Sydney, Australia (a revised version is available on the Web from the authors).

This is a pioneer application of sentiment analysis in the political sphere. The technique is based on machine learning, and pragmatic information (about who is speaking to whom and whether or not they are expressing agreement) is included.

Having reviewed the key literature, we move on to describing our approach and our new software tool.

4.0 OUR APPROACH TO AUTOMATING SENTIMENT

We took two paths to automating sentiment. First, we pushed the frontier on BoWs analysis. Second, we broke ground in a new direction by mining utterances and collecting information on speech acts. To do so we developed a new software tool we call Pathos (meaning s*entiment* in Greek). Below we describe the new software tool.

4.1 Pathos

Pathos performs three tasks. It classifies documents, performs BoWs analysis, and measures utterances and speech acts. Figure 2 illustrates our three pronged approach.

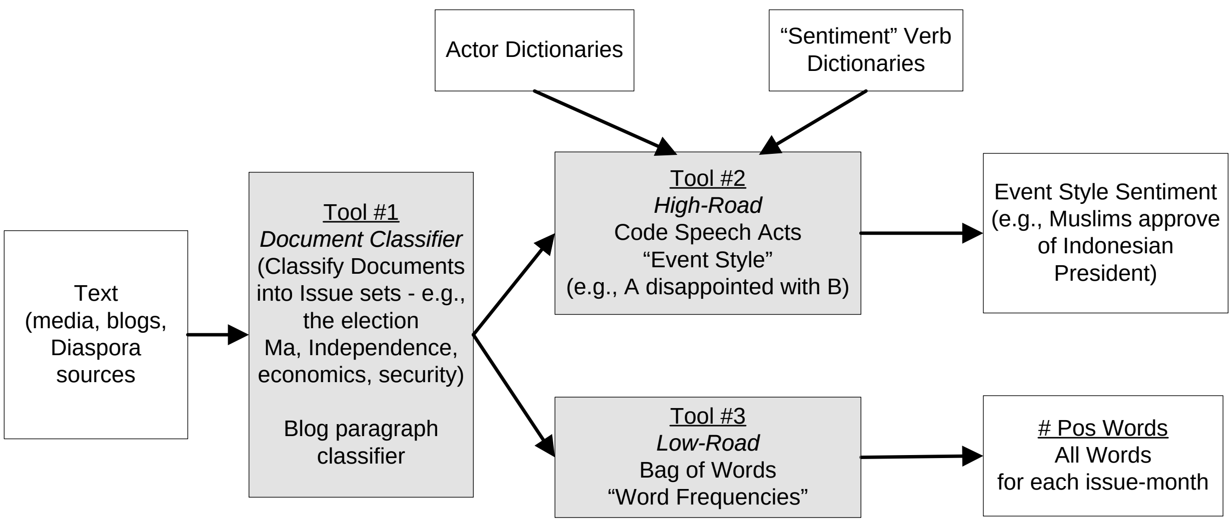


Figure 2: Pathos Software Tool

4.1.1 Text Classification. One way to isolate the target of sentiment is to classify the documents into issue-oriented categories. Prior to the classification, Pathos' document classifier was trained with a set of documents relating to the respective category. As mentioned above, after consultation with Sean O'Brien at DARPA we focused our analysis on Taiwan. Each news article and blog posting was classified into one of three categories: Ma Ying-Jeou (the most recently elected President), Security/Cross Strait Relations (between Taiwan and China), or the Economy. These categories were chosen because they related to the subject matter of the Taiwanese media publications and polling questions.

Our prototype text classifier uses a "vector similarity" approach. This is a well-known technique introduced by Salton, Wong, and Yang (1975). A word frequency table is constructed for each document; all the tables list the same words in the same order. Tables of the frequencies of *n* words are treated as vectors in *n*-dimensional space and compared by calculating the cosine of the angle between the vectors. This is used as a goodness-of-fit measure ranging from 0 to 1 and is insensitive to the length of the document; the size of the document affects the length of the vector but not its direction.

Each training set (collection of documents of known class) is treated as a single large document, and each of the documents to be classified is compared to all of the training sets. Each document is assigned to the class to which it has the highest goodness-of-fit. Our software lemmatizes all words (reduces to dictionary form) so that, for instance, *elect, elected,* and *electing* are grouped together (but not *election* or *electable,* which would have separate entries in a dictionary). Words are not weighted for importance because the vector comparison method already gives more weight to words that are more effective in distinguishing document classes. After texts are classified they are fed into the BoWs or the Speech Acts analyzer.

4.1.2 BoWs. BoWs analysis corresponds to the "low road" method of capturing sentiment. (A "BoWs" is a text viewed simply as words each occurring a certain number of times, but ignoring context.) A list of words indicating positive and negative sentiment was obtained from publications of the Harvard General Inquirer project (http://www.wjh.harvard.edu/~inquirer/). After adding more words to each list, each word was rated by a collection of linguists on a scale from negative one to one based on their implied sentiment. For instance, words such as excellent and great received values close to one, while words such as despair and revolt received values close to negative one.

First, each news article or posting is classified into one of the three categories previously mentioned. Pathos then calculates Polarity, PolarityNZ, PolarityW, SubjectivityNZ, Subjectivity, and Splitness (defined below) of each document based on these hand-scaled values. For example, take the sentence, "I think he is good and you think he is great but she thinks he is bad." Assuming the polarities of *good, great,* and bad are +0.5, +1.0, and -0.5 respectively, and the other words are not considered to indicate sentiment, this translates into:

0 0 0 0+.5 0 0 0 0 0 +1 0 0 0 0 0 -.5

We then calculate the following measures:

- Polarity, [(2 × sum of positive numbers in list)/(sum of absolute values of all numbers in list)] – 1 or [(2*1.5)/2]-1=.5 (see Godbole et al., (2007))
- PolarityNZ, polarity/number non-zero words, .5/3 = 0.1667
- PolarityW, polarity/number words, .5/17 = 0.0294
- SubjectivityNZ, sum of |each word|/number non-zero words, 2/3 = 0.6667
- Subjectivity, sum of |each word|/number words, 2/17 = 0.1176
- Splitness, Subjectivity-Abs(PolW), 0.1176 - 0.0294 = .0882

The polarity measures refer to ways of measuring the overall positive or negative tone of a text, while the subjectivity measures tap the overall strength of sentiment. Splitness refers to how much contradiction occurs within a text. Inconsistent texts have higher splitness scores.

With regard to additional measures of polarity, we also tried summing the numbers and dividing by non-zero words (.33) in earlier iterations but those values were highly correlated with polarity). At the end of the day, all of these measures are highly correlated and provide similar results. The one we use most often is the one that the literature uses most often, the Godbole et al., (2007) polarity measure. The BoWs analysis is beneficial in that it will track the overall "mood" of the media on a certain subject/topic.

4.1.3 Utterances/Speech Acts. In the early 1990s, the Kansas Events Data System (KEDS) demonstrated that the collection of events data could be automated (Schrodt & Gerner, 1994; Schrodt, Davis, & Weddle, 1994). With automated coding, the coding rules are

transparent, the data are easily and quickly reproducible, the data can be regenerated using alternative coding schemes, and the data are unaffected by individual coders' biases, as well as reducing the time required for coding from hundreds of hours of human labor to mere minutes once the input texts have been formatted and coding dictionaries prepared. This has radically changed the information that is available to conflict scholars. Moreover, the KEDS project has spawned a number of similar projects, and this technology has spilled over into a variety of other areas of political science as well.

KEDS and its open-source successor, TABARI program[1] were originally used to collect information primarily on regional interactions among actors (e.g., the Levant). TABARI uses a "sparse-parsing" technique to extract the subject, verb, and object from a sentence and determines the appropriate codes using pattern matching on actor and verb dictionaries.[2] The result is a numeric representation of an event in the form of "someone does something to someone else" on a certain day.

Pathos's utterances/speech acts coder captures sentiment in an event style format; that is, with the source, target, and verb (expressing or reporting the sentiment) individually identified. The primary difference from events data is that only sentiment verbs (verbs identified as those conveying sentiment) are used. As such a new sentiment verb dictionary had to be created. This new dictionary has over 800 verbs and verb phrases. Each of these sentiment verbs was rated on a scale from negative one to one (similar to the method for BoWs). Actors making the statements were includes in an actor dictionary developed specifically for Taiwan. Targets included individuals included in the new Taiwan actor dictionary as well as terms focusing on the economy, countries, organizations, and security. There were over 200 Taiwanese-specific actors alone. The advantage of speech acts is that one can observe who is directing sentiment towards whom, and any events/political implications that sentiment might produce.

To code such utterances and speech acts, Pathos performs part-of-speech tagging on all input, using the Penn Treebank tag set and a lexicon derived from the Penn Treebank (http://www.cis.upenn.edu/~treebank/). The tagger is a hand-optimized Brill tagger (Brill, 1995). For Pathos, the tagger has been tested on material representative of this project, and numerous small improvements have been made. The most important are:

- The rules for distinguishing between verb past participle (VBN) and verb past tense (VBD) have been refined, leading to more accurate identification of active and passive forms.
- The non-Penn tag verb past participle, active (VBNA) has been added. This tag replaces VBN when preceded by a form of *have.* Its effect is to peel off the active-voice VBN forms so that only the passive-voice verbs in the text are tagged VBN.
- Numerous proper nouns (NNP) were tagged as common nouns (NN) by the Penn Treebank. Accordingly, rules to correct this, taking capitalization into account, have been added.

[1] See http://raven.cc.ukans.edu/~keds/index.html for information on the KEDS and TABARI projects. Also see Schrodt (1996; 2006) for the respective codebooks.

[2] TABARI recognizes pronouns and dereferences them. It also recognizes conjunctions and converts passive voice to active voice (Schrodt 1998).

The tagger in Pathos also performs lemmatization, i.e., guided by part-of-speech tags, it reduces each word to its “dictionary form,” such as *having* → *have, children* → *child*, etc. Lemmatization differs from stemming in the following way: Stemmers just chop off endings (*having* → *hav*), but lemmatizers produce actual words with nearly 100% correctness (which is not hard to achieve if the tags are correct). Tagging and lemmatization improves our ability to code utterances and speech acts accurately. Pathos differs from TABARI in these and other ways.

Pathos implements a TABARI-like event-coding function, but internally, the pattern matching mechanism is different:

- Pathos works on lists of words which have pre-computed attributes such as tags and lemmas, whereas TABARI treats the input as a character string;
- Pathos performs lemmatization rather than (or as an alternative to) what TABARI calls stemming (i.e., prefix matching);
- Pathos has considerably richer resources for syntactic disambiguation because tagging has been performed.

As a result, Pathos codes sentiment utterances and speech acts more accurately than TABARI.

There were two methods used to create speech acts data. The first uses an actor and sentiment verb dictionary to code speech acts. Then, the data is filtered according to the target of the speech act. For instance, if the target “Ma” was chosen, only speech acts in which sentiment is directed towards Ma Ying-Jeou will have been kept.

The second method of creating speech acts data is slightly different. First, each document is classified into one of the aforementioned categories. Pathos then searches each document for any speech acts. Unlike the first method of creating speech acts data, only targets defined as part of the government are kept, regardless of classification. In this method, the government is being isolated as the object of sentiment so as to create a proxy for hard to capture targets. This method was implemented for models with documents classified as “the Economy” and “Security/Cross Strait Relations.”

5.0 RESEACRH DESIGN

We designed our analysis with two main goals in mind. First, we wanted to demonstrate internal validity or the degree to which our measures correlated with similar measures (polling data). Second, we wanted to demonstrate that our measures were externally valid – that they correlated with other measures that they should be correlated with (e.g., measures of the economy, political performance, etc.). To achieve these objectives we collected additional data to compare our measures to as well as generated our own sentiment measures using Pathos. We not only correlated our measures with polling data and measures of the economy, we also used our new measures in models of politics and compared results across models which included our measures to those which excluded our measures. To eliminate the possibility of generating results as an artifact of (1) the unit in which we chose to analyze our data or (2) a specific polling question, we aggregated our data into both weekly and monthly temporal units and examined the relationship between our measures and multiple polling questions. Below we elaborate on our data sources.

5.1 Data

All data were gathered for the time period April 2007 through July 2008 to capture the period of time leading up to and just after a prominent Taiwanese Presidential election. We collected news reports, blog texts, polling data, events data, and other country-level data such as economic measures to complete our analysis.

5.1.1 News Sources. In order to perform the analysis, over 1 GB of text was downloaded. The following Taiwanese news publications were used for this analysis: China Post, Tapei Times, Central News Agency, Kuomintang (KMT) News Network, and Taiwan Review. These account for the most popular sources of general Taiwan news available in English.

5.1.2 Blog Sources. Among the blogs downloaded include Forumosa: Taiwan Politics, That's Impossible: Politics from Taiwan, Far Eastern Sweet Potato, Taiwan Matters, Sun Bin, Rank, Only Red Head in Taiwan, Jerome F. Keating's Writings, and It's Not a Democracy it's a Conspiracy. These were found to be the most popular and most accessible blogs emanating from Taiwan.

5.1.3 Polls. Polls were gathered from a variety of poll-administrating agencies namely United Daily News, United Evening News, Global Views Magazine, TVBS News, China Times, Taiwan Apple Daily News, ERA Television, and The Executive Yuan Research, Development and Evaluation Commission. There were hundreds of questions administered, however only 33 were used for this analysis. Question topics included the economy, culture, international relations, support for various leaders within the government, and many more. That said, the polling data are very spotty in the sense that they do not occur with any regular frequency. Many of the polling questions are asked with months and weeks in between. Around the election, the questions regarding Ma become more frequent.

5.1.4 Events Data. We also collected events data of who was doing what to whom in addition to our sentiment measures. Ultimately we wanted to relate our sentiment measures to

the on goings of Taiwan politics. We used TABARI to independently collect the events data for our analysis. Doing so would eliminate any possible claims that Pathos biased our findings.

5.1.5. Other Country Level Data. We also collected information on monthly levels of consumer prices, unemployment, and Gross Domestic Product (GDP). These data were obtained from the Taiwan Ministry of Finance.

5.1.6 Data Aggregation. We ran the news sources and the blog sources through our document classifier, BoWs analyzer, and speech acts analyzer as described above. The next step was to take the raw output and aggregate it into usable metrics in analyses. We essentially chose to average our data across both weekly and monthly intervals so as not to generate findings only relevant to one unit of temporal aggregation. Shellman (2004) shows that how we aggregate our data can impact the inferences we draw from models of politics. Shellman suggests that we temporally aggregate our data into multiple units and run the same models. Results which hold across different units should be given more weight. We designed our research such that our results could not be attributable to the way in which we chose to aggregate our data.

As noted above, the polling data we were using to compare our measures are reported irregularly and infrequently. So comparing our measures to polls often include only 15-20 data points. This is an asset of our new measures; our automated sentiment measures can be calculated at a daily level. Nevertheless, to demonstrate the validity of such measures we needed to compare them to alternative measures generally respected in the academic and policy communities. Polling data more often than not reflect three-day rolling averages of responses to the survey instrument.[3] Thus, we calculated three-day moving averages from our daily automated sentiment measures to reflect how polls report their data.

5.2 Methodology

We employ several methods to analyze the utility of our new sentiment data. They can best be divided up into methods which test the internal and external validity of our new measures.

5.2.1 Internal Validity Methodology. To demonstrate the internal validity of our sentiment measures we compared our measures to Taiwan polling data (three day averages). We focused on two questions pertaining to Ma (the most recent elected President), two questions focusing on the economy and two questions focusing on security issues – specifically relations with China. Our analyses for internal comparisons focused on computing standard bivariate Pearson correlation coefficients.

5.2.2 External Validity Methodology. To demonstrate external validity, we correlated our sentiment measures with consumer prices, inflation, unemployment, and GDP. We also generated events data using TABARI and specified vector autoregression (VAR) models to examine the relationships over time between actors' political actions and sentiment towards them. VAR models are econometric models used to capture the interdependence among multiple time series. The model essentially specifies an equation for each time series as a function of lags of itself plus lags of the other time series in the model. Including lags of the dependent variable biases against finding support for our hypotheses since lagged dependent variables tend to soak up much of the variance in a regression model. For example, if we wanted to know if Taiwanese

[3] For example, see http://www.gallup.com/poll/109897/gallup-daily-obama-moves-ahead-48-42.aspx.

sentiment affected Taiwan's actions towards China and we also hypothesized that China's actions also impacted Taiwan's actions, we would specify a VAR model that estimated Taiwan's actions as a function of Taiwan's past actions towards China, China's past actions towards Taiwan, and Taiwanese Sentiment directed by the Taiwan citizens towards Taiwan. To further explore whether or not sentiment mattered and to what degree in such a model of Taiwanese actions towards China, we would run a model which only included China's actions towards Taiwan and compare it to the model which included lags of our sentiment measure. Finally, we would compute Granger causality statistics to tell us if sentiment was indeed having an impact independently of Taiwan's prior actions toward China and China's prior actions toward Taiwan. A Granger causality test essentially determines whether one time series is useful in forecasting another. For example, X is said to Granger cause Y if lags of X provide statistically significant information about future values of Y in the presence of lagged values of Y. That is, does X provide additional information which improves the forecast of Y. We report both Granger causality tests and the comparative model statistics to show how sentiment impacts political behavior and how political behavior impacts sentiment.

6.0 RESULTS

We split our results into internal and external validity but also by the unit of analysis (weekly v. monthly) and the polling question. We performed much more analysis than can be summed up and reported here. For the purposes of this report, we focus on the most pertinent results though all of our analyses were reported and presented to DARPA on April 18, 2009 via power point format.

6.1 Interpreting the Graphs

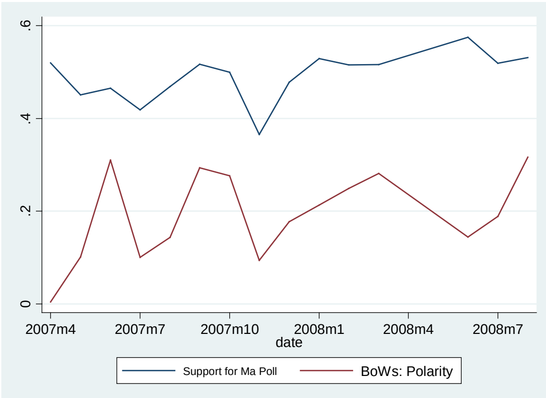
Before we move on to the results, we want to explain our graphs and how to interpret them. There are two types of graphs presented below. The first type plots the automated sentiment measures (BoWs and Speech Acts) against the polling data. The polling data are always displayed as blue lines and the sentiment measures are displayed as red lines. We also report the correlation coefficients (r) below each graph to show the strength of the relationship. The correlation coefficients show how the series move together. We display the results in graphic form in addition to the correlation coefficients because it is often the case that the lines are moving together in the general direction or are at generally the same level, yet the correlation coefficients are not as strong as one would think because the coefficients examine each point rather than general trends. Graphic displays allow us to view such general trends.

The second type of graph we display is our model fit. To show the fit of our models we plot the actual series of the dependent variable (blue) against the model fitted or predicted values (red). Essentially, the model predicted values are the values produced from the model. The more the red overlaps with the blue, the better the model fits the data. We also report the correlation coefficients (r) between the actual and predicted/fitted series.

6.2 Ma Monthly Results

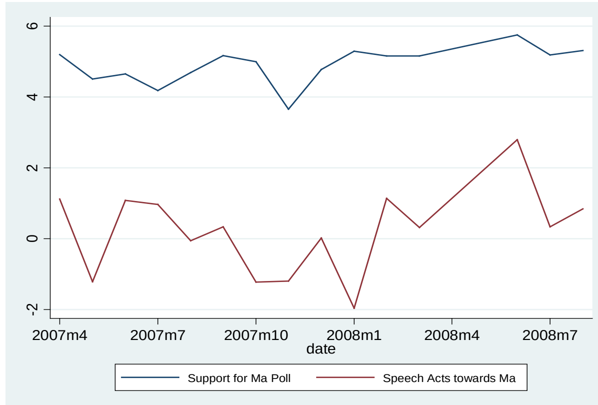
We first correlated our BoWs and speech acts measures with the polling question that asked "Who do you support for President – Ma, Hseih, or undecided?" We used the percentages for support and correlated them with our BoWs and Speech Acts measures. We display the graphs for the measures we calculated from all of our texts (media reports and blogs), but we also report the correlation coefficients calculated between the polls and our measures calculated using only the media reports and only the blogs. The results are displayed in Figure 3.

A. BoWs Polarity to Polls (Support)



r = .35

B. Speech Acts to Polls (Support)



r = .42

Figure 3: Comparisons of Monthly Ma Series

Figure 3A shows the relationship between the BoWs polarity measure and the polling data we collected on “support for Ma.” Figure 3B shows the relationship between the speech acts measure and the same polling data we collected on “support for Ma.” The correlations coefficients are .35 and .42, respectively. While both measures seem to track the polling data, our speech acts measure yields a higher correlation to the polling data. This is a consistent finding across all of our results; the speech acts measures more closely track the polls than their BoWs counterparts. This is also the case when we separate our media reports from our blogs in Table 1. Speech acts calculated from just media reports correlate with the poll data at .44 while the BoWs measures only correlate at .21. Yet, blogs, generally typed by one person with one voice, are opposite. In fact, the BoWs blog measures correlate at .21 while the speech acts calculated from blogs correlate negatively with the polling data. We’ll revisit this finding once we have discussed the results for the weekly data as well as the security and economic data below.

Table 1: Breakdown Across Media & Blogs (Months)

	BoWs	Speech Acts
All Media & Blogs	.35	.42
All Media	.21	.44
All Blogs	.41	-.60

Figure 4 shows our ability to model and therefore forecast the polling data from our sentiment data. We put the polling data on the left-hand side of our model as the dependent variable and included lags of our sentiment measures on the right-hand side as independent variables. Figure 4 plots the actual polling data against the predicted polling data generated from our sentiment model. As one can see, the model fitted values closely resemble the actual polling data. The two series correlate at .92. As such, if policy-makers really want polling data, we could use our data to generate it for them.

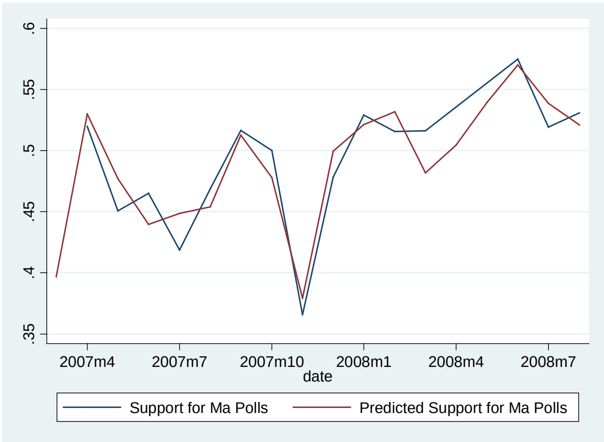


Figure 4: Model of Polls Using Our Sentiment Measures (Months)

Table 2 reports the results of our initial external validity analyses. We correlated the polling data as well as our sentiment measures to the monthly consumer price index, a measure of monthly inflation, and a quarterly GDP measure obtained from the Ministry of Finance. The overall results show that our speech acts measure correlates in the same direction to all the external economic measures as well as the polling measures. This demonstrates the external validity of our measure. The BoWs measure performs similarly with respect to monthly consumer prices but correlates in the opposite direction with inflation and GDP than the polling and speech acts indicators. Again, our speech acts measure seems to outperform our BoWs measure on external validity criteria.

Table 2: Correlating Ma Sentiment Measures to Economic Measures (Months)

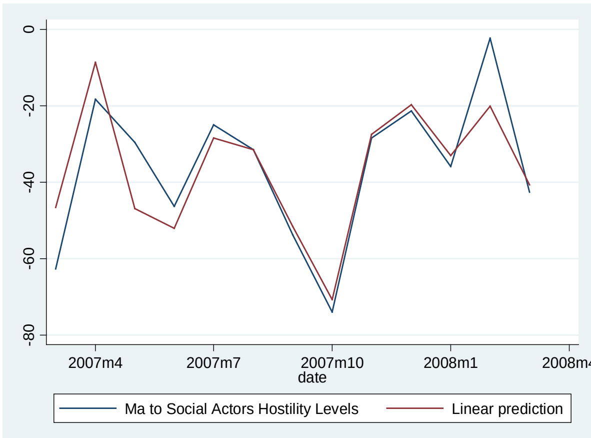
	Ma Polls (Support)	Ma Speech Acts	Ma Polarity (BoWs)
Monthly CPI	.49	.24	.19
Monthly Inflation	.04	.04	-.63
Quarterly GDP	-.26	-.37	.21

The next analysis we performed included our sentiment measures in models of Ma's hostile and cooperative actions. We generated events data using TABARI as discussed above and isolated the actions taken by Ma towards social actors (citizens, labor unions, etc.). We attached Conflict and Mediation Event Observations (CAMEO) weights[4] ranging from -10 to +10 to each of the actions. We then summed all of the negative events over each month to create a composite indicator of Ma's hostile actions and we summed all of the positive events over each month to create a composite indicator of Ma's cooperative actions. We then calculated our monthly speech act measures by isolating social actors and Ma as the target of the utterances. In sum, we now had a measure of Ma's hostile and cooperative actions towards social actors and a measure of social actors "sentiment" towards Ma.

Figure 5 shows the results of three different VAR models focusing on Ma's *hostile* actions as the dependent variable of interest: (A) models which include all three variables, (B) models which include only the sentiment variables, and (C) models which exclude the sentiment variables. Part D of Figure 5 reports the Granger causality test statistics. Figure 5A shows that the model fitted values (red) closely resemble Ma's actual hostile actions (blue). Figure 5B shows that a model of his hostile actions which only includes lags of sentiment is actually a good fitting model emphasizing the explanatory power of our speech acts indicator. Figure 5C shows the model fitted values of a model which excludes our speech acts indicator. As one can see in Figure 5C, excluding our sentiment data from the model results in a weak fitting model yielding less explanatory power. Our model containing only sentiment data is a better fit than the model which excludes such information. The Granger causality test statistics reported in Figure 5D confirm our other results. All of these series Granger cause each other meaning that all of the series are related to each other and add to explaining variance in each other.

[4] See http://web.ku.edu/~keds/papers.dir/ISA08.pdf for more information on the CAMEO coding scheme. See http://web.ku.edu/~keds/cameo.dir/CAMEO.SCALE.txt for the CAMEO scale values.

A. Complete VAR Model of Hostile Actions



r = .89

B. VAR Model of Hostile Actions using only Sentiment

r =.59

C. VAR Model of Hostile Actions without Sentiment

r = .48

D. Granger Causality Tests

	Hostility
Hostility	51.08***
Cooperation	24.43***
Sentiment	15.39***

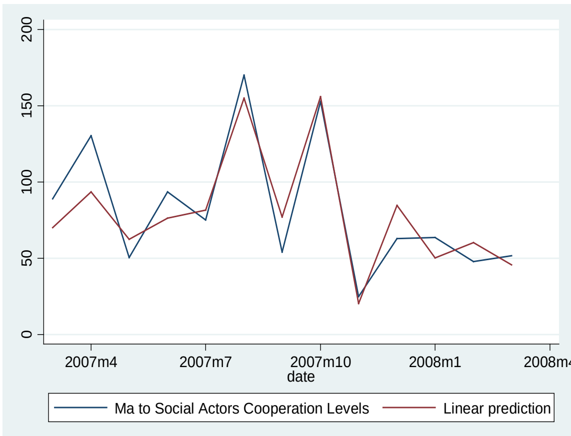
*** = statistically significant at the .01 level

Figure 5: VAR Models of Ma Hostile Actions (Monthly)

Figure 6 shows the results of three different VAR models focusing on Ma’s *cooperative* actions as the dependent variable of interest: (A) models which include all three variables, (B) models which include only the sentiment variables, and (C) models which exclude the sentiment variables. Figure 6D reports the Granger causality test statistics. Figure 6A shows that the model fitted values (red) closely resemble Ma’s actual cooperative actions (blue). Like Figure 5B, Figure 6B shows that a model of his cooperative actions which only includes lags of sentiment is also good fitting model. Again, this emphasizes the explanatory power of our speech acts indicator. Figure 5C shows the model fitted values of a model which excludes our speech acts indicator. As one can see in Figure 6C, excluding our sentiment data from the model results in a weak fitting model yielding less explanatory power. Our model containing only sentiment data provides just as good of fit as the model which excludes such information. The Granger causality

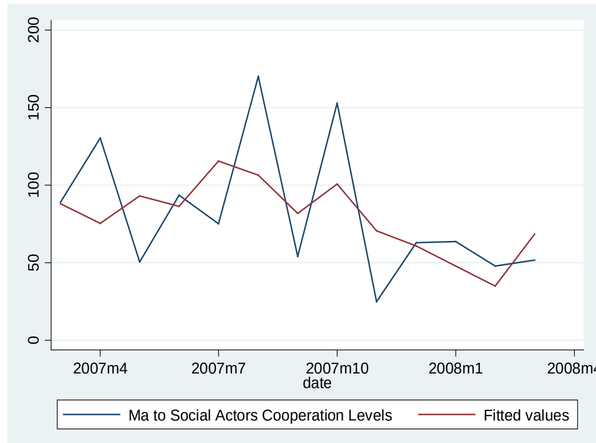
test statistics reported in Figure 6D confirm our other results. All of these series Granger cause each other meaning that all of the series are related to each other and add to explaining variance in each other.

A. Complete VAR Model of Cooperative Actions



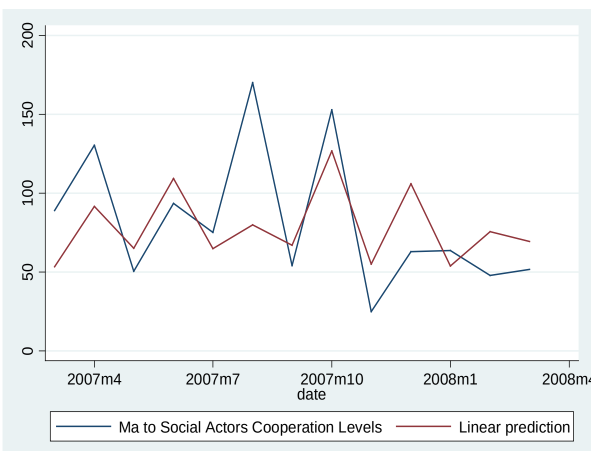
R = .89

B. VAR Model of Cooperative Actions using only Sentiment



R =.53

C. VAR Model of Cooperative Actions without Sentiment



R = .55

D. Granger Causality Tests

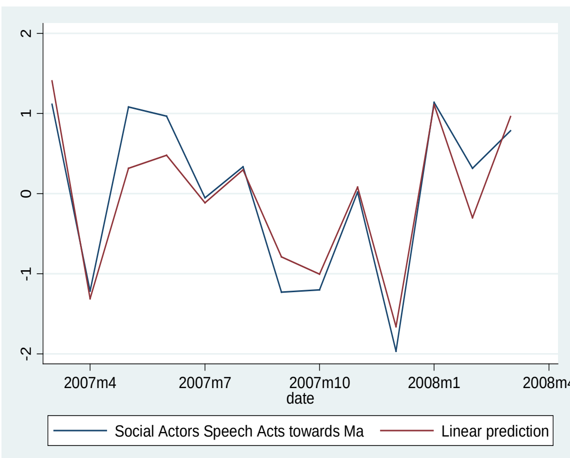
	Cooperation
Hostility	9.44**
Cooperation	42.89***
Sentiment	29.66***

Figure 6: VAR Models of Ma Cooperative Actions (Monthly)

Figure 5 and Figure 6 together reveal the importance of including sentiment in our models of politics. The analyses confirm our hypothesis that politicians respond to public attitudes and that such attitudes shape political behavior. Figure 5C reveals specification error and omitted variable bias in models that exclude measures of attitudes and sentiment which unfortunately are most models of political behavior.

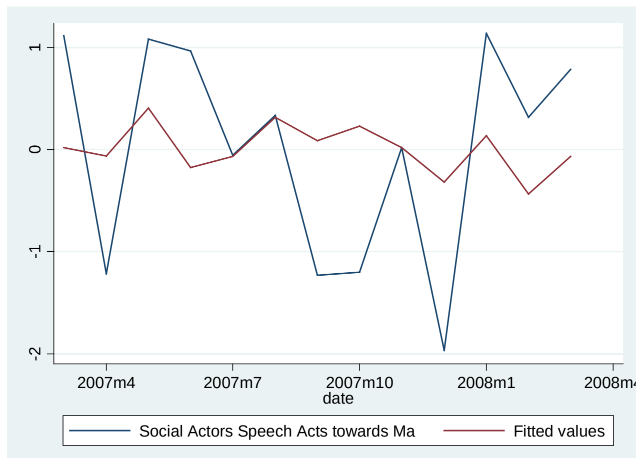
Figure 7 shows the results of how Ma's hostile and cooperative actions affect levels of sentiment. Figure 7A shows how closely the model fitted values follow the actual sentiment data we collected. However, when we take Ma's hostile and cooperative actions out of the model, we observe that our sentiment model falls apart. Such results indicate that political sentiment is driven by the actions of politicians. The Granger causality statistics reported in 7C also confirm our conclusions. To be sure, this is not a novel insight, yet given that our model reflects such results yields confidence in our data and our automated measures. In other words, the fact that our model produces such a result lends credence to our software tool and our abilities to measure sentiment daily from electronic texts.

A. Complete VAR Model of Sentiment



R = .92

B. VAR Model of Sentiment without Ma Hostility and Cooperation



R =.22

C. Granger Causality Tests

	Sentiment
Hostility	14.31***
Cooperation	34.10***
Sentiment	56.39***

Figure 7: VAR Models of Sentiment (Monthly)

6.3 Ma Weekly Results

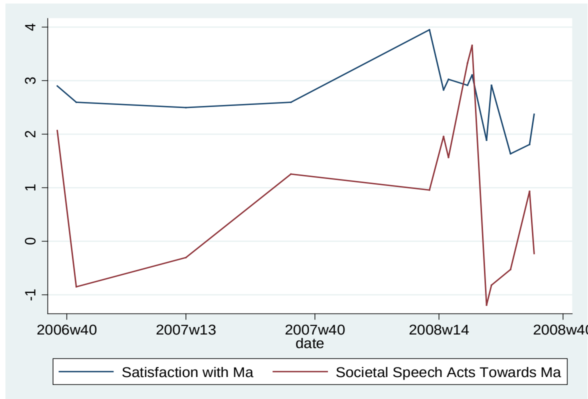
We performed identical analyses as those reported above on weekly aggregated data so that our conclusions were not an artifact of the unit of aggregation we chose. We also focused on a different polling question collected by a different agency so as not to allow our findings to be an artifact of the polling question or firm we selected. In short, our additional analyses produce very similar inferences.

Figure 8 compares our sentiment measures to the polling data at weekly intervals. Again, our speech acts indicator correlates at a higher level than our BoWs indicator. Table 3 breaks down the weekly series into our measures calculated solely from media reports and solely from blogs. Speech acts outperform the BoWs measures across all sources at the weekly level.

A. BoWs Polarity to Polls (Support)

R = .35

B. Speech Acts to Polls (Support)



R = .48

Figure 8: Comparison of Weekly Ma Series

Table 3: Breakdown Across Media & Blogs (Weekly)

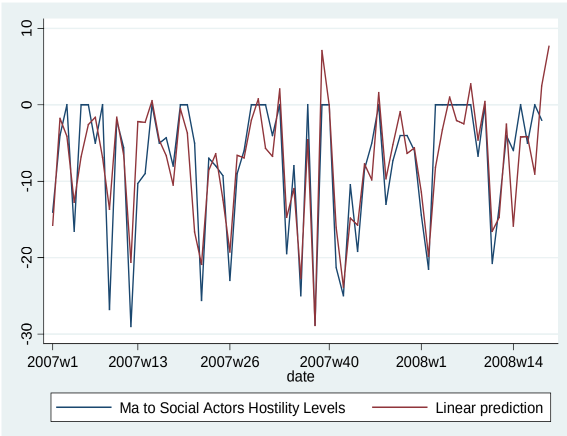
	BoWs	Speech Acts
All Media & Blogs	.35	.48
All Media	.32	.46
All Blogs	.47	.50

Similarly, we compared our weekly measures and the polling data to external indicators (not depicted in a table for reasons of redundancy). Like the monthly level, our Speech acts measures were correlated in the same direction as the polling measures with external indicators such as Consumer Price Index (CPI), inflation, and unemployment. For example, unemployment was correlated with the polling data at -.67 and with the speech acts indicator at -.51. Such

correlation coefficients reveal that as unemployment goes up, support for Ma goes down. Again, this is a finding to expect. However, being able to show it with our automated data provides credibility to our software tool and our sentiment measures.

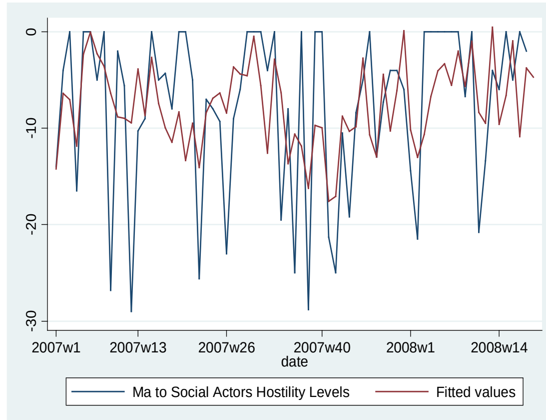
Like Figure 5, Figure 9 shows the results of three different VAR models focusing on Ma's *hostile* actions as the dependent variable of interest: (A) models which include all three variables, (B) models which include only the sentiment variables, and (C) models which exclude the sentiment variables. Figure 9D reports the Granger causality test statistics. Figure 9 of course focuses on observations aggregated at the weekly rather than the monthly level. Figure 9A shows that the model fitted values (red) closely resemble (.86) Ma's actual hostile actions (blue). Figure 9B shows that a model of his hostile actions which only includes lags of sentiment is a decent fitting model emphasizing the explanatory power of our speech acts indicator alone. Figure 5C shows the model fitted values of a model which excludes our speech acts indicator. Excluding our sentiment data from the model results in a weaker model than when it is included in the model. The Granger causality test statistics reported in Figure 9D reveal that all of these series Granger cause each other meaning that all of the series are related to each other and add to explaining variance in each other. Figure 9A also reports the correlation coefficient for the model run using the BoWs indicator rather than the speech acts indicator. While the actual and predicted values correlate at .86 using the speech acts indicator, the series only correlate at .47 using the BoWs polarity measure in the same model specification. This suggests that our speech acts indicator is a more robust measure of social actors' sentiment. This supports our contention that measuring speech acts is a better way to measure mass sentiment than a BoWs technique.

A. Complete VAR Model of Cooperative Actions



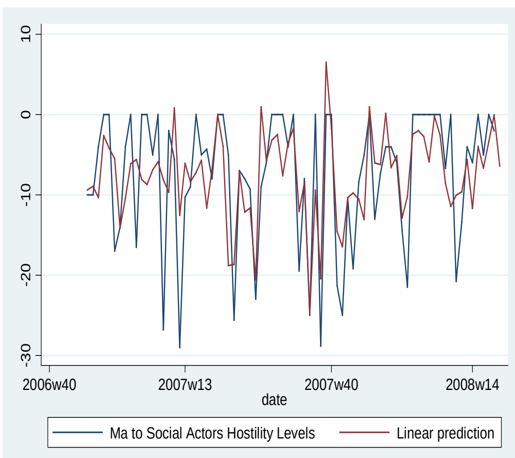
r = .86; BoWs r = .47

B. VAR Model of Cooperative Actions using only Sentiment



r =.50

C. VAR Model of Cooperative Actions without Sentiment



r = .66

D. Granger Causality Tests

	Cooperation
Hostility	9.44**
Cooperation	42.89***
Sentiment	29.66***

Figure 9: VAR Models of Ma Hostile Actions (Weekly)

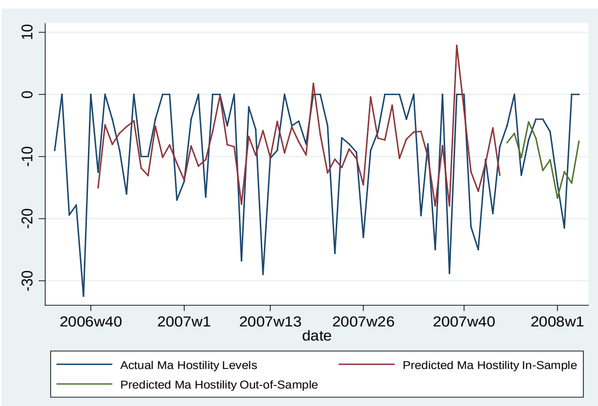
In addition to modeling Ma's hostile actions we also modeled his cooperative actions at the weekly level. While we don't report the results here given their monotony, they reflect the same patterns and trends as the monthly results. The full VAR model predicted values are correlated with the actual values at .82. When the BoWs measure replaces the speech acts indicator in the model, the actual and predicted series only correlate at .46, again revealing that the speech acts indicator is superior to the BoWs indicator. The model which excludes sentiment produces actual and predicted series correlated at.55, while the model including only sentiment produces actual and predicted correlated series at .42. Finally, the Granger causality results confirm that all the series Granger cause each other.

Finally, we also examined how Ma's weekly hostile and cooperative actions impacted sentiment. The full VAR model reveals that the actual and model predicted values of sentiment

are correlated at .88. Moreover, the model of sentiment which includes only lags of itself produces a correlation coefficient of only .56 between the actual and model predicted series. The results confirm that weekly sentiment is driven by Ma's hostile and cooperative actions much like monthly sentiment.

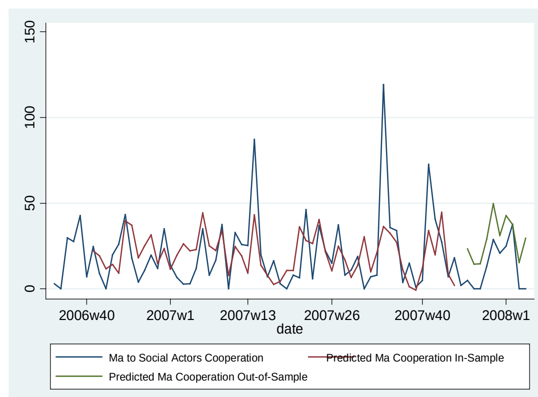
One of the important capabilities that decision-makers show interest in are models' abilities to forecast behavior. As such, we used our model to forecast Ma's behavior out of sample and compare our forecasts to the actual data we collected. We used our weekly model to do so and display our forecasts 10-weeks ahead of Ma's hostile and cooperative actions in Figure 10A and 10B, respectively. Each of the graphs shows that our model forecasts trend well with the actual actions Ma takes. Our cooperative action forecasts in Figure 10B correlate with his actual actions at .88, while our hostile action forecasts correlate with his actual hostile actions at .59. Our models even forecast his hostile actions and cooperative actions out 25 weeks and those predictions correlated with his actual hostile and cooperative actions at .41 and .64, respectively. Obviously, forecasts are more prone to error the farther in time one predicts actions. However, our models were able to forecast the trends in Ma's behavior fairly accurately even 25 weeks (6 months) ahead.

A. Ma's Hostility



r-out-of-sample = .59

B. Ma's Cooperation



r-out-of-sample =.88

Figure 10: Out-of-Sample Forecasting (10 Weeks Out)

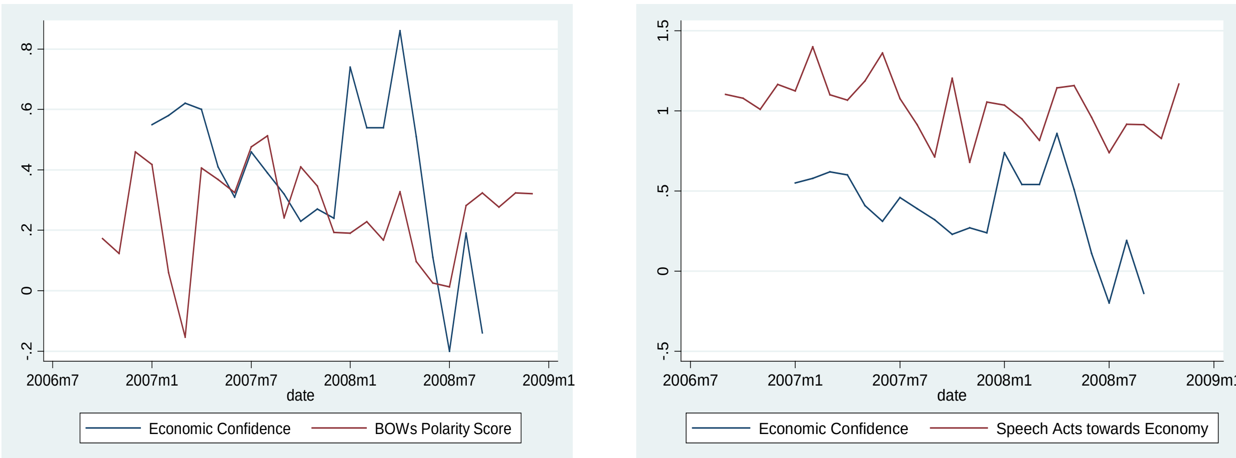
6.4 Monthly Economic Results

In addition to examining how sentiment impacts a leader's actions, we also wanted to explore how sentiment was related to the economy. Thus, we first compared our measures of sentiment to survey data on economic confidence and then we used our measures of sentiment to examine how attitudes towards economic performance affected Ma's actions. We report our findings for our monthly temporal unit below.

We first correlated our BoWs and speech acts measures with the polling question that asked "how much confidence do you have in the economy?" We display the graphs for the measures we calculated from all of our media reports because there were not enough blog postings on the economy to include them in our analyses. Given the different scales in the series

we linearly transformed them so we could depict them on the same graph. The results are displayed in Figure 11.

A. BoWs Polarity to Polls (Economic Confidence) **B. Speech Acts to Polls (Economic Confidence)**



r = .05 r = .42

Figure 11: Comparisons of Monthly Economic Series (Linearly Transformed for Scale purposes)

One can see that our speech acts measure yet again outperforms our BoWs measure in terms of its ability to track well with the polling data. The speech acts measure captures the overall trends in the data well. Moreover, when we used the lags of our speech acts indicator to model the economic polling data, we observed actual and model predicted values correlated at .90. Again, we can reproduce the polling data fairly accurately from the speech acts data.

Table 4 reports the results of our initial economic external validity analyses. We correlated the economic confidence polling data as well as our sentiment measures to the monthly consumer price index, monthly inflation levels, and a monthly unemployment measure obtained from the Ministry of Finance in Taiwan. The overall results show that our speech acts measure correlates in the same direction to all the external economic measures as the polling measures. Our speech act measure even correlates at roughly the same level as the economic confidence measure to the CPI indicator. As expected, as prices go up, confidence and sentiment goes down. The same relationship holds for unemployment. These results demonstrate the external validity of our measure. The BoWs measure performs similarly with respect to monthly consumer prices but correlates in the opposite direction with inflation and unemployment than the polling and speech acts indicators. Again, our speech acts measure seems to outperform our BoWs measure on our initial external validity criteria.

Table 4: Correlating Economic Sentiment Measures to Economic Measures (Monthly)

	Economic Confidence Poll	Speech Acts	Polarity (BoWs)
Monthly CPI	-.57	-.50	-.22
Monthly Inflation	.29	.08	-.26
Unemployment	-.71	-.47	.09

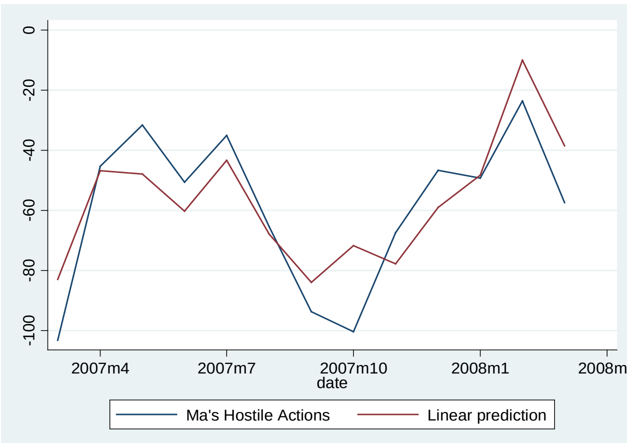
The next analysis we performed included our sentiment measures in models of Ma's hostile and cooperative actions. Some literature in political science (e.g., Gilpin, 1987) suggests that the economy drives politics. Moreover, a subset of literature argues that economic approval drives politicians' decisions. As such, we developed a model of Ma's political actions which were driven, in part, by economic approval. To do so, we generated events data using TABARI as discussed above and isolated the actions taken by Ma towards social actors (citizens, labor unions, etc.). We attached CAMEO weights[5] ranging from -10 to +10 to each of the actions. We then summed all of the negative events over each month to create a composite indicator of Ma's hostile actions and we summed all of the positive events over each month to create a composite indicator of Ma's cooperative actions. We then calculated our monthly speech act measures by isolating social actors as the actors and the economy as the target of the social actors' utterances. In sum, we now had a measure of Ma's hostile and cooperative actions towards social actors and a measure of social actors "sentiment" towards the economy. In short, we wanted to examine if attitudes about the economy impacted Ma's political behavior.

Figure 12 shows the results of three different VAR models focusing on Ma's *hostile* actions as the dependent variable of interest: (A) models which include all three variables, (B) models which include only the economic sentiment variables, and (C) models which exclude the sentiment variables. Figure 12D reports the Granger causality test statistics. Figure 12A shows that the model fitted values (red) closely resemble Ma's actual hostile actions (blue). The two series correlate at .80. Figure 12B shows that a model of his hostile actions which only includes lags of economic sentiment is actually a good fitting model emphasizing the explanatory power of our speech acts indicator. The actual and predicted series correlate at .68, only a .12 difference from the full model's values. Figure 12C shows the model fitted values of a model which excludes our speech acts indicator. As one can see in Figure 12C, excluding our economic sentiment data from the model generates a weaker fitting model. Our model containing only sentiment data is a better fit than the model which excludes such information. The Granger causality test statistics reported in Figure 12D confirm our other results. Like our previous results, all of these series Granger cause each other meaning that all of the series are related to each other and add to explaining variance in each other. The results for Ma's cooperative actions

[5] See http://web.ku.edu/~keds/papers.dir/ISA08.pdf for more information on the CAMEO coding scheme. See http://web.ku.edu/~keds/cameo.dir/CAMEO.SCALE.txt for the CAMEO scale values.

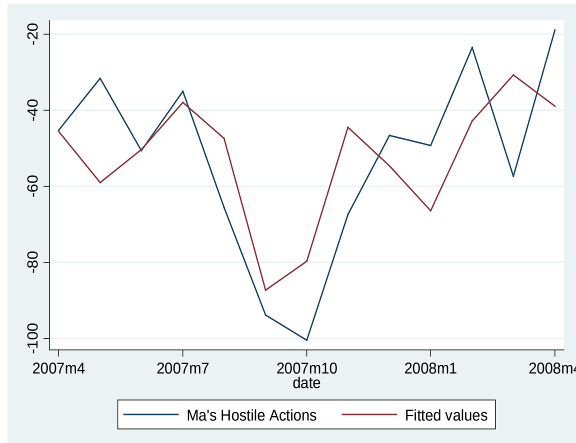
were very similar and revealed the same patterns and inferences. In short, our economic speech acts variable plays an integral role in improving the model fit of Ma's hostile and cooperative actions. At the same time, our results not depicted here, also show that economic sentiment is explained by Ma's actions.

A. Complete VAR Model of Hostile Actions



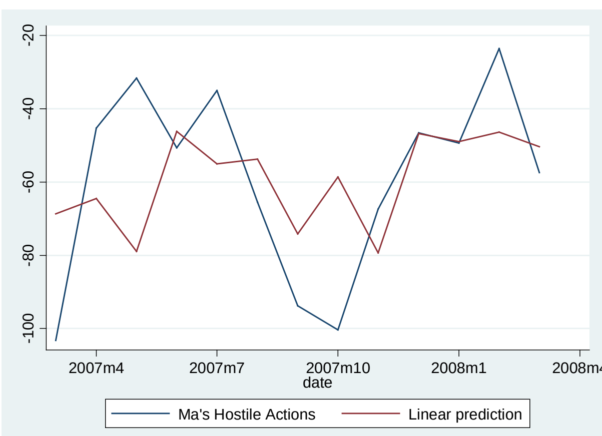
r = .80

B. VAR Model of Hostile Actions using only Sentiment



r =.68

C. VAR Model of Hostile Actions without Sentiment



r = .44

D. Granger Causality Tests

	Hostility
Hostility	21.86**
Cooperation	16.51***
Sentiment	12.04***

Figure 12: VAR Models of Ma Hostile Actions (Weekly)

The analyses once again confirm our hypothesis that politicians respond to public attitudes and that such attitudes shape political behavior. Figure 12C further reveals specification error and omitted variable bias in models that exclude measures of attitudes and sentiment. Once more, this is not a novel finding, but the fact that our measures are performing in these models in the ways we articulated above provides credibility to our software tool and ability to generate valid and reliable automated measures of sentiment.

6.5 Weekly Security Results

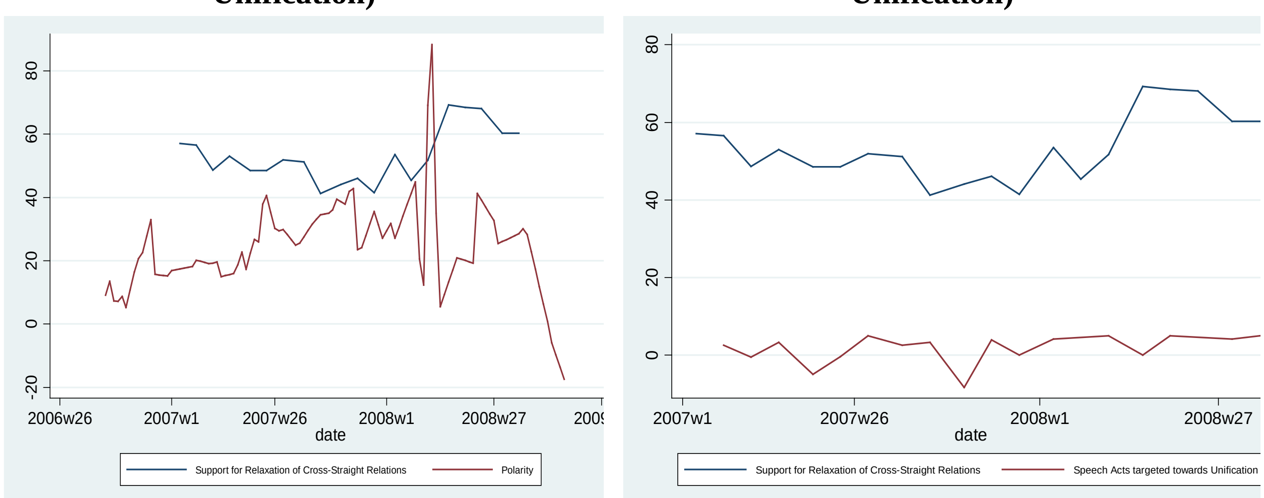
In addition to modeling the effects of sentiment towards Ma and towards the economy on Ma's actions, we also wanted to examine the impact of sentiment on Taiwan-China cross-strait relations. There is a social science literature which stresses the impact of domestic attitudes and audiences on how governments behave abroad. While Taiwan is considered to be a part of China, the territory conducts foreign relations with many sovereign states although *de facto* relations are conducted with nearly all other states. Moreover, Taiwan has expressed desire to secede from the mainland at times throughout the course of history and the independence/unification issue is always a main cleavage in which elections are fought over. For our analyses, we concentrated on Taiwanese attitudes and opinions towards independence/unification.

We first classified electronic documents into our category that dealt with cross-strait relations. We then performed a BoWs analysis on such documents. For our speech acts, we focused on unification and independence as the targets of utterances. The polling question we analyzed asked respondents about their support for unification. Thus, we matched our speech act measure that focused on unification as the target of sentiment to the unification polling question data.

We display the graphs for the measures we calculated from all of our texts (media reports and blogs), but we also report the correlation coefficients calculated between the polls and our measures calculated using only the media reports and only the blogs. The results are displayed in Figure 13.

A. BoWs Polarity to Polls (Support for Unification)

B. Speech Acts to Polls (Support for Unification)



R = -.32

R = .38

Figure 13: Comparisons of Weekly Security Series

Figure 13a shows the relationship between the BoWs polarity measure and the polling data we collected on "support for unification." As to be expected, our BoWs measure performed horribly. Given that documents could be about both unification and independence and the BoWs approach (the standard in the literature) has no way to separate support for one or the other, we

expected muddy results. We in fact found a negative correlation between our BoWs measure and the polling data on unification.

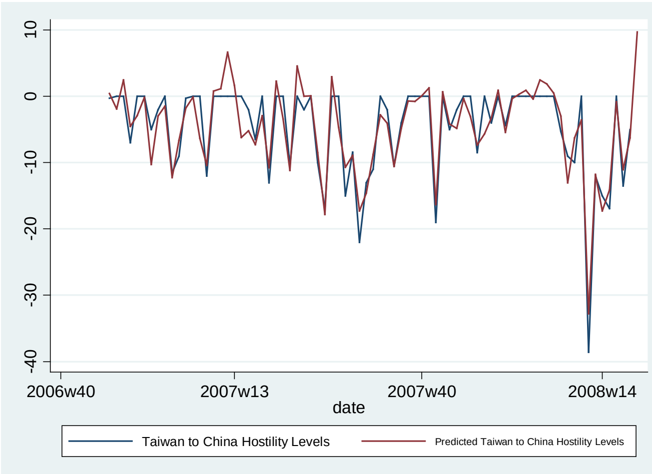
Figure 13b shows the relationship between the speech acts measure focusing on unification and the same polling data. The correlation coefficient is .38 and the graph shows that the series tend to trend together in more less the same way. The speech acts measures displayed in Table 5 broken up into just media reports and just blogs more closely track the polls than their BoWs counterparts. In fact, the speech acts calculated from blogs are more highly correlated with the unification polling data than speech acts pulled out of both the media reports and the blogs combined.

Table 5: Breakdown of Unification Across Media & Blogs (Weekly)

	BoWs	Speech Acts
All Media & Blogs	-.32	.38
All Media	-.46	.32
All Blogs	.19	.42

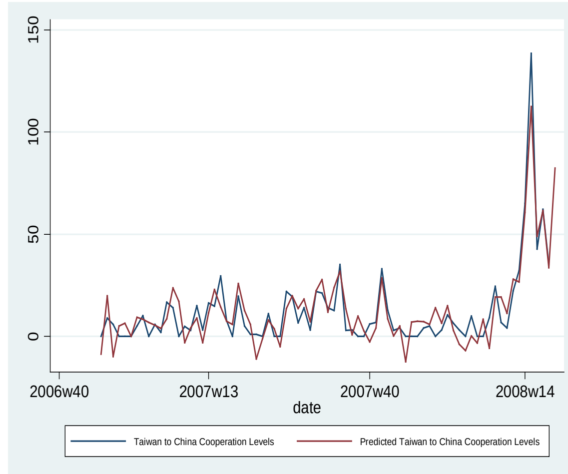
The results of the primary analysis we wanted to perform with these data are reported in Figure 14. We first generated events data using TABARI as discussed above and isolated the actions taken by Taiwan towards China and China towards Taiwan. We split these actions into both hostile and cooperative actions, and then we added our weighted speech acts towards unification to the mix. We then estimated a VAR model.

A. Taiwan to China Hostility



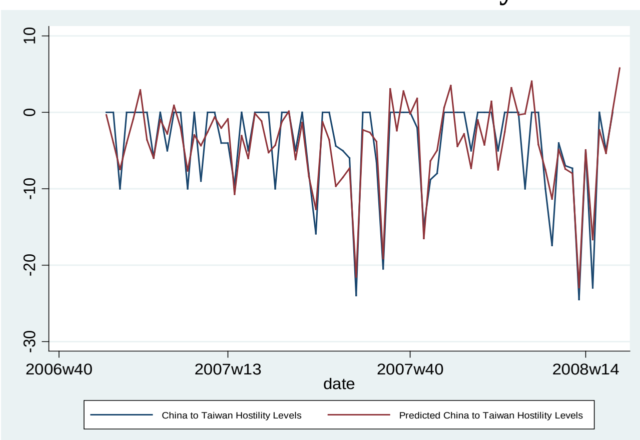
r = .92; .62 w/o sentiment

B. Taiwan to China Cooperation



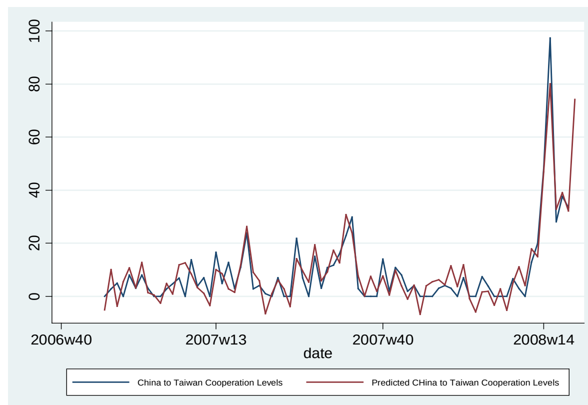
r =.92; ,81 w/o sentiment

C. China to Taiwan Hostility



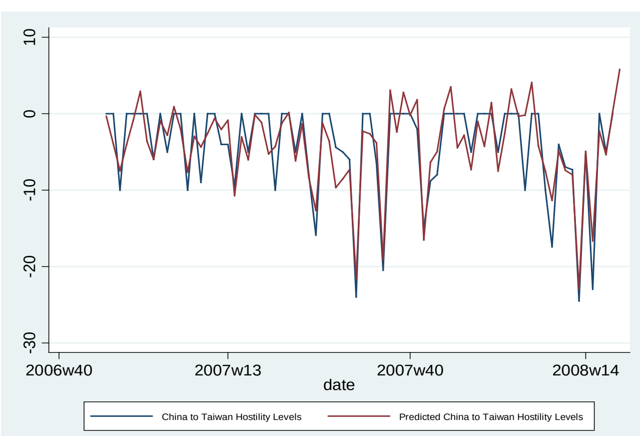
r = .44; .65 w/o sentiment

D. China to Taiwan Cooperation



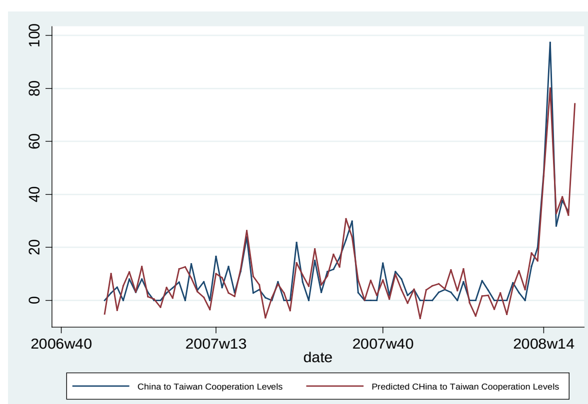
r=.93; .79 w/o sentiment

E. Speech Acts



r = .93

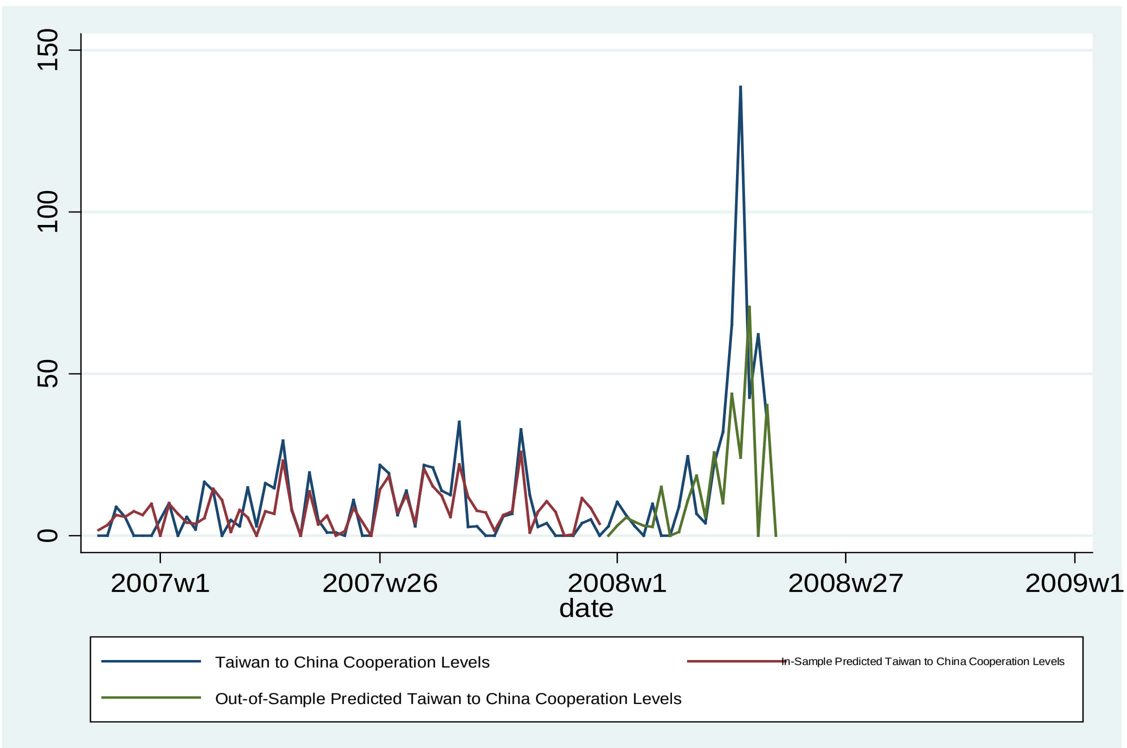
F. Taiwan to China Cooperation as a Function of Sentiment Only



r=.60

Figure 14: VAR Models of Taiwan-China Relations (Weeks)

Figure 14 shows the results of VAR models focusing on China-Taiwan relations. To begin, our model of Taiwan to China hostility generates predicted values that correlate with the actual values at .92 (see Figure 14A). When one removes our speech act variable from the model, the correlation between actual and predicted values falls to .62. The same pattern is present in terms of China to Taiwan hostility. The model including our speech acts measure results in predicted and actual values correlated at .89, while the model that removes sentiment from the model yields a correlation coefficient between the actual and predicted values of .65 (see Figure 14C). Figure 14B and 14D show similar patterns when examining Taiwan to China and China to Taiwan cooperation levels. The full model which includes our measure of sentiment provides a better fit to the data than the model which excludes our sentiment measure. Figure 14E shows that the model also explains speech acts well.[6] Finally, Figure 14F illustrates how well sentiment alone explains Taiwan to China cooperation levels. Finally, Figure 15 shows how well our model performs in forecasting Taiwan-China cooperation levels. Figure 15 shows a 20 week forecast in green and the out-of-sample predicted values correlate with the actual values at .50. Moreover, visually, one can see that the model forecasts well the general trends in Taiwan-China cooperation levels.



In-r = .77; Out-r = .50

Figure 15: Out of Sample Forecast from FULL VAR Foreign Policy Model Taiwan Cooperation Towards China (20 weeks out)

[6] We should note that all of the Granger causality tests are also statistically significant revealing that all of these series are interrelated and aid in explaining variance in each other.

Taken together, the results illustrate the interrelationships among sentiment and actions and, in particular, the importance of sentiment in explaining political actions. The analyses confirm our hypothesis that cross-strait relations are, in part, affected by domestic attitudes.

6.6 DARPA SAE ICEWS Rebellion Model with Sentiment as an Added Factor

The final analysis we report here included our measure of sentiment in a previously developed model for the DARPA Integrated Crisis Early Warning System (ICEWS) project that Strategic Analysis Enterprises, Inc. (SAE) worked on in 2007-2008. Our model previously aimed at forecasting rebellion as defined by the DARPA ICEWS program using events data generated from TABARI and other structural indicators. The original model focused on 29 countries in South and Southeast Asia. However, given our resources we focused on Australia, Indonesia, and the Philippines and generated our sentiment measure from our available media reports we had garnered already for the ICEWS effort. We then ran our rebellion model without sentiment on these data which covered these three countries from 1998-2004 and forecasted predictions for 2005-06. Next, we included our sentiment indicator in the same model and again ran the model on the 1998-2004 data and forecasted predictions for 2005-06. The results for our sentiment model are displayed in Table 6.

Table 6: ICEWS Rebellion Model with Sentiment Included

Variable	Coef	SE	t-score	p-value
Social Actor Sentiment Towards Government	-0.15*	0.093	-3.77	0.005
Social Actor Sentiment Towards Dissidents	0.31*	0.180	2.29	0.035
Government Hostility towards Separatists	0.49***	0.104	4.71	0.000
Government Hostility towards Social & Religious Actors	0.43*	0.235	1.82	0.069
Separatists Material Cooperation	1.79*	0.973	1.83	0.067
Separatists Low Hostility	2.09***	0.680	3.07	0.002
Separatists High Hostility	0.68***	0.260	2.62	0.009
Constant	-3.17	0.49	-6.43	0
In Sample Accuracy	97%			
Out Of Sample Accuracy	90%			

Table 6 illustrates that our social actor sentiment towards the government indicator is negative and statistically significant at the .10 level. In other words, when sentiment is positive towards the government, the probability of rebellion is reduced. Alternatively, our social actor sentiment towards dissidents measure is positive and statistically significant in our model. This indicates that as sentiment grows more and more positive towards the dissidents, the probability of rebellion increases. This is an empirical analysis of the hearts and minds argument. Specifically, when more people side with the government the likelihood for separatist rebellion is

lowered. Moreover, when more people side with the dissidents, the probability of separatist rebellion is higher. Moreover, our model provides out of sample accuracy over 90%.

First our model demonstrates that winning hearts and minds is an important variable in explaining the probability of rebellion. Second, if we can measure how governments are doing at winning hearts and minds, we can better predict when we might observe rebellions. Our method of collecting sentiment could be employed to gauge the opinions of the masses on the ground and indicate on any given day the levels of support for the government and the dissidents. As a result, we could measure how we are doing and how other governments are doing at winning hearts and minds. This could be a way of understanding how well our strategic messaging campaigns are performing, how well our infrastructure development is being received, and/or how well a variety of our actions or the actions of other government and/or non-government actors are being received. Until now, such analyses have not been possible on a daily basis, using automated methods of extraction. This is a significant achievement.

7.0 FUTURE TECHNICAL IMPROVEMENTS

We believe that our software has great potential to play a large role in future state of the art analyses of political conflict. As such, we recently committed to invest another $100,000 into Pathos, our software tool, and additional dictionary development to improve upon our initial prototype. Below we discuss each of these plans for future improvement.

7.1 SAE Internal Research and Development (IRAD)

While we have proven our prototype produces the data we desire for analysis of sentiment and its endogenous effects on political conflict, quality research always leads to new questions, new desires, and a new wish list of capabilities.

7.1.1 Verb Dictionary. The verb dictionary used to code speech acts consists of only verbs that convey sentiment (e.g. support, condemn, approve). The expansion of this sentiment verb dictionary allows for the recognition of more speech acts. There are currently over 2,300 verbs or verb phrases in the sentiment verb dictionary. We plan to expand this under our IRAD project. Moreover, our dictionary was primarily based on the CAMEO coding scheme developed by Phil Schrodt and limited to sentiment verbs only. We have found the scheme to be limiting for both event coding and sentiment coding. Specifically, the direct objects of the verbs are attached to the root verbs. This can create long lists of verb phrases and at times inaccurate placement of verbs under specific codes. For example, the root verb "said" is followed by a long list of various things an actor can say and they are all given the same CAMEO code. In reality, there are many things one can say and various sayings should be weighted differently. We broke some of these key verb phrases up in our DARPA seedling but much more needs to be done to the verb dictionary for more accurate coding of utterances and speech acts. We believe this additional verb dictionary work alone can improve coding 50-70%. We will overhaul the entire scheme breaking up each verb from its direct object, lemmatizing each verb, and then creating a new direct object dictionary. At the end of the day, our coding effort will now generate an extra piece of information. In the past our scheme generated the date, actor, target, and verb. Following our IRAD work, our scheme will generate the date, actor, target, verb, and direct object. Each direct object will be independently weighted so for example "Bill said John is an idiot" is given a different code than "Bill said John is smart." This should greatly improve our coding accuracy.

7.1.2 Actor Dictionaries. Pathos identifies unknown (i.e. not in the actor dictionary) actors as it codes a set of text. This allows for limitless expansion of the actor dictionaries. As more actors are added to the dictionary, the software can not only code speech acts, but the sources and targets that correspond to a given speech act. Our actor dictionary for this effort concentrated on Taiwan, but we could employ our technology anywhere in the world and allow Pathos to build new dictionaries based on its abilities to find actors not already in an actor dictionary. Moreover, we currently have extensive dictionaries already built for more than 30 countries.

7.1.3 Pathos Improvements. We also plan to make several improvements to Pathos during our IRAD effort.

Performance, reliability, and ease of use

- Reorganize hastily written code, especially in the user interface
- Use a profiler to find slow-moving code sections and speed them up
- Simplify and speed up the pattern matching algorithm
- Improve user interface ergonomics
- Provide some kind of “project file” to define an entire work session
- Save user choices and have them be defaults for next run

Input format

- Add ability to split a long input file at lines that match a string or regular expression
- Integrate improvements to the date parser
- Implement a date-sensitive actor dictionary format
- Allow varying the names of the dictionary and input data files

Text classification module

- Use TF*IDF or modified TF*IDF instead of simple vector classification
- Implement unsupervised classification (clustering) as well as existing supervised classification system

Event/Speech Act coding module

- Improve pronoun coreferencing
- Add pattern matching options: match by lemma, by tag; skip limited numbers of words
- Numerical Referencing: The ability to differentiate between the phrases: “students plant 450 bombs at schools” and “students plant bombs at schools.” Pathos nor any other state of the art software accounts for the number ‘450’ whatsoever in the given sentence.

At the end of our IRAD, we should have a software program ready for action on major projects and ready for deployment in multiple areas of research.

8.0 FUTURE PROJECTS

Our software program and analysis of its output illustrate its utility for national and international security analysis. While we have shown sentiment data is a necessity for understanding the ebb and flow of political conflict, there are myriad studies such data can be applied to that would yield increased understanding of phenomena. We specifically focus on strategic communications, messaging, and effects-based operations analyses below.

8.1 Strategic Communications

Strategic communication means essentially getting the right message, through the right media, to the right audience at the right time and with the right effect. Our sentiment analysis tool can help us understand if we are achieving each of these goals. Ultimately, we want to know if our message reached the targeted audience and what effects it has. We can control for each of these variables (timing, media, message, audience, and effect) allowing one or two to vary in experiments and analyze whether or not for example, our message is having the desired effect. One way forward would be to match cases on these variables and examine the impact of the message on the effects. Alternatively we could match cases on the message and other various factors, and alter the timing component and observe its effects. No matter how one slices it, we must be able to understand how the message is received and how such receipt of the message affects actions. Our sentiment analysis tool can collect this necessary information for the analysis in near-real time.

8.2 Effects Based Operations

Another aspect of national security analysis that sentiment analysis plays a direct role in is “effects based operations” (EBO) planning. EBO requires planning, executing and assessing operations to attain the effects required to achieve desired national security objectives. In essence, EBOs model the adversary as a system as opposed to a single actor. Such models require monitoring and emphasizing direct, indirect, and complex effects among variables in a system of systems. Models highlight cumulative and cascading effects in which time and space must be considered. Sentiment is a key intervening variable in that US actions will affect various populations of social and political actors and their attitudes will affect how they respond to such actions. Following our actions, targeted actors’ sentiment precedes our adversary’s reactions, and influences such reactions. Our tool and analyses can examine how our actions affect sentiment and how such sentiment affects our adversary’s actions. In many ways, our tool and data generated from our tool can help us know if and how we are “winning hearts and minds.”

9.0 CONCLUSION

In this seedling, we developed a prototype software tool to generate sentiment data from electronic text documents (specifically media reports and blogs). Overall, our project was successful in that we demonstrated that the data generated from our new tool were both internally and externally valid indicators of sentiment. Our data closely mirrored public opinion data obtained from various Taiwan polls and correlated in the same direction as the polling data with other external economic indicators such as consumer prices and unemployment. We further demonstrated that our data were useful in understanding and forecasting political actions of politicians (e.g., President Ma) as well as the dyadic interactions of governments (i.e., Taiwan-China relations). Our project overall has shown that we can create much needed and valuable sentiment data faster, better, and cheaper than polling data. Polls are not always able to be taken in every environment (e.g., poor conflict-ridden countries and locations) and when completed often cost a large sum of money for translators, survey administrators, and survey designers. Finally, more often than not they are difficult to carry out every day. Our product can fundamentally change the way we do business.

Extensions of our project can yield new insights into why specific actions do not often yield the intended consequences. Such theoretical and technical steps will yield more effective and accurate evaluations of effects-based operations, strategic communications, and subsequently more precise forecasts of the effects of specific government activities and actions.

REFERENCES

Brill, Eric. “Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging.” *Computational Linguistics* **21(4)**, 1995, 543-565.

Bueno de Mesquita, B., Morrow, J., Siverson, R.M., & Smith, A. "An institutional explanation of the Democratic peace." *American Political Science Review*, **93(4)**, 1999, 791-807.

Gilpin, R. *The political economy of international relations.* Princeton, NJ: Princeton University Press, 1987.

Godbole, N., Srinivasaiah, M., & Skiena, S. “Large-scale sentiment analysis for news and blogs.” *ICWSM*, Boulder, CO., 2007. Available online at http://icwsm.org/papers/3--Godbole-Srinivasaiah-Skiena.pdf

Guevara, C. "General principles of guerrilla warfare." In B. Loveman & T.M. Davies Jr. (Eds.), *Guerrilla warfare*. Lincoln, NE: University of Nebraska Press, 1985, 47-69.

Salton, G., Wong, A., & Yang, C.S. “A vector space model for automatic indexing.” *Communications of the ACM*, **18(11)**, 1975, 613-620. Available on line at http://www.cs.uiuc.edu/class/fa05/cs511/Spring05/other_papers/p613-salton.pdf

Schrodt, PA., & Gerner, D.G. “Validity assessment of a machine-coded event data set for the Middle East, 1982-92.” *American Journal of Political Science*, **38(3)**, 1994, 825-854.

Schrodt, P.A., Davis, S.G., & Weddle, J.L. “Political science: KEDS – A program for the machine coding of event data.” *Social Science Computer Review*, **12(4)**, 1994, 561-587.

Shellman, S.M. “Time series intervals and statistical inference: The effects of temporal aggregation on event data analysis.” *Political Analysis*, **12(1)**, 2004, 97-104.

Tse-Tung, M. *Basic tactics*, (S.R. Schram, translator). New York: Praeger, 1966.

LIST OF ACRONYMS

BoWs	Bag of Words
CAMERO	Conflict and Mediation Event Observations
CPI	Consumer Price Index
DARPA	Defense Advanced Research Projects Agency
DRT	Discourse Representation Theory
EBO	Effects Based Operations
GDP	Gross Domestic Product
ICEWS	Integrated Crisis Early Warning System
IRAD	Internal Research and Development
KEDS	Kansas Events Data Systems
KMT	Kuomintang
NN	Common Nouns
NNP	Numerous Proper Nouns
SAE	Strategic Analysis Enterprises, Inc.
SMEs	Subject Matter Experts
SSA	Social Science Automation
TABARI	Text Analysis by Augmented Replacement Instructions
VAR	Vector Autoregression
VBD	Verb Past Tense
VBN	Verb Past Participle
VBNA	Verb Past Participle, Active

GLOSSARY

1. Sentiment - All reflections of support, liking, opposition, or disliking of individuals/groups, their actions and/or proposals, and events.
2. Bag of Words (BoWs) - unordered collection of words, disregarding grammar and even word order.
3. Blog - Website, usually maintained by an individual with regular entries of commentary, descriptions of events, or other material such as graphics or video.
4. Polling Data - Answers to survey questions given to a sample of individuals. Our data come from various firms in Taiwan geared toward Taiwan politics.
5. Speech Acts - Coded sentiment verbs.
6. Posting - Similar to an article in news sources, it is the smallest single unit of analysis within a blog.
7. Polarity - The overall positive or negative tone of the text (without regard to what is being said about whom).
8. PolarityNZ - Polarity per non zero word
9. Nonzero words – words that express positive or negative tone (e.g., good, best, bad, worst, etc.)
10. Zero words – words that do not express positive or negative sentiment (e.g., a, the, he, she, went, travelled, etc.)
11. PolarityW - Polarity per total number of words
12. SubjectivityNZ - Similar to PolNZ but using the absolute values of the numbers to measure overall strength of sentiment.
13. Subjectivity - Similar to PolW but using the absolute values of the numbers to measure overall strength of sentiment.
14. Splitness - A measure of how much contradiction there is within the sentiment expressed in the text. Inconsistent texts have higher splitness.
15. Sentiment Verbs - Verbs identified as those conveying sentiment
16. Locution - In linguistics, it is what the authors says.
17. Illocution - In linguistics, it is what the author means.
18. Perlocution - In linguistics, it is the effect of the author’s expression.
19. Directed dyad - A pair of actors (e.g., A and B) in which one actor directs an action, expression, or behavior towards the second actor. A to B is different from B to A.